Widespread recombination within human parechoviruses: analysis of temporal dynamics and constraints

K. S. M. Benschop, Ç. H. Williams, K. C. Wolthers, G. Stanway and P. Simmonds

Correspondence
K. S. M. Benschop
k.s.benschop@amc.uva.nl

1Department of Medical Microbiology, Laboratory of Clinical Virology, Academic Medical Center, Amsterdam, The Netherlands
2Department of Biological Sciences, University of Essex, Colchester CO4 3SQ, UK
3Virus Evolution Group, Centre for Infectious Diseases, University of Edinburgh, Edinburgh EH9 1AJ, UK

Human parechoviruses (HPeVs), members of the family Picornaviridae, are classified into six types. To investigate the dynamics and likelihood of recombination among HPeVs, we compared phylogenies of two distant regions (VP1 and 3Dpol) of 37 HPeV isolates (types 1 and 3–5) and prototype sequences (types 1–6). Evidence for frequent recombination between HPeV1, 4, 5 and 6 was found. The likelihood of recombination was correlated with the degree of VP1 divergence and differences in isolation dates, both indicative of evolutionary times of divergence. These temporal dynamics were found to be most similar to those of human enterovirus species B variants. In contrast, HPeV3 remained phylogenetically distinct from other types throughout the genome. As HPeV3 is equally divergent in nucleotide sequence from the other HPeV types, its genetic isolation may reflect different biology and changed cellular tropisms, arising from the deletion of the RGD motif, and likely use of a non-integrin receptor.
Lukashev, 2005; Oberste *et al.*, 2004a, b; Santti *et al.*, 1999; Simmonds & Welch, 2006).

To study the likelihood and dynamics of recombination within parechoviruses, we analysed several HPeV sequences obtained from two distant regions within the genome (VP1 and 3Dpol). In total, 37 HPeV isolates were obtained from the Netherlands (n=29), California (n=6) and Finland (n=2). A full listing of the HPeV sequences used is available as Supplementary Table S1 in JGV Online. Twenty-nine samples were sequenced and typed previously based on the VP1 region and submitted to GenBank under the accession numbers DQ172416, DQ172418, DQ172420–DQ172421, DQ172424–DQ172428, DQ172430, DQ172432–DQ172433, DQ172435–DQ172446, DQ172448, DQ172451 (Dutch isolates containing six-digit numbers, prefixed with NL; Benschop *et al.*, 2006a), AM234724, AM234726 and AM234728 (Californian isolates; Al-Sunaidi *et al.*, 2007; Schnurr *et al.*, 1996). The remaining samples were typed within the VP1 region as described previously (Al-Sunaidi *et al.*, 2007; Benschop *et al.*, 2006b). To determine partial 3Dpol sequences of each HPeV isolate, RNA was extracted by using a QIAamp Viral RNA mini kit according to the manufacturer’s instructions (Qiagen). The extracted RNA was reverse-transcribed and amplified by using a SuperScript III One-Step RT-PCR kit (Invitrogen) according to the manufacturer’s instructions. The following primers were designed within the conserved region within the 3Dpol sequence by aligning all known HPeV types: OL1502 (5’-GTGTACAGGATGATCATGATGGA-3’) and OL1501 (5’-CTTAGTCAAACACCATGGGCA-3’, nt 7253–7233), where nucleotide numbering is relative to that of the HPeV1 Harris strain (GenBank accession no. S45208). Amplicons were isolated by agarose-gel electrophoresis and purified by using a QIAquick gel extraction kit according to the manufacturer’s instructions (Qiagen), using a spin-column protocol. DNA was eluted in 50 µl water. BigDye Terminator sequencing reactions were performed by GeneService, Cambridge, UK (http://www.geneservice.co.uk).

The HPeV VP1 and 3Dpol sequences were aligned by using CLUSTAL W (Thompson *et al.*, 1994) and edited manually by using the SIMMONICS sequence editor (version 1.6; http://www2.warwick.ac.uk/fac/sci/bio/research/devans/bioinformatics/simmonics/). Neighbour-joining phylogenetic trees based on the nucleotide sequence were constructed separately for the VP1 and 3Dpol regions of the HPeV genome by using the MEGA 3.1 software package (Kumar *et al.*, 2004) with Jukes–Cantor (J-C)-corrected distances (Jukes & Cantor, 1969) (Fig. 1). Sequence data were bootstrap-resampled 1000 times to determine robustness of the observed grouping; branches supported by >70% of replicate trees are indicated. All available full-length sequences for HPeV were obtained from GenBank and were included in the analysis: HPeV1 strains Harris (S45208) and BNI-788St (EF051629); HPeV2 strain Williamson (AJ005695); HPeV3 strains A308-99 (AB084913) and Can82853-01 (AJ889918); HPeV4 strains K251176-02 (DQ315670) and T75-4077 (AM235750); HPeV5 strains CT86-6760 (AF055846) and T92-15 (AM235749); HPeV6 strains NII561-2000 (AB252582) and BNI-67/03 (EU024629). All HPeV sequences were monophyletic and formed phylogenetic groups that corresponded to their type designation in the VP1 region (609 nt) (Fig. 1). However, within the 3Dpol region (706 nt), a breakdown of the type-specific segregation was observed; sequences in this region of HPeV1, 4, 5 and 6 frequently failed to group according to type. The majority of recently isolated HPeV1 strains showed 3Dpol sequences distinct from that of the prototype HPeV1 strain Harris, first isolated in 1956. As an initial indication of the time-related nature of the recombination events in HPeV, isolates that clustered closely together in the VP1 region (pairwise-compared J-C distances of <0.0601) almost invariably remained together in the 3Dpol region. When a greater VP1 divergence between HPeV1 isolates was observed, a loss of segregation was consistently observed in the 3Dpol region. These findings are consistent with the previously observed incongruent phylogenetic relationships within the non-structural region of HPeV4 and 5 (Al-Sunaidi *et al.*, 2007; Benschop *et al.*, 2006b).

To test formally whether the loss of segregation between variants within each of the parechovirus types was related to their degree of evolutionary and epidemiological separation (as indicated by their divergence in VP1), members of the same HPeV type were classified further by their bootstrap-supported phylogenetic groupings within the 3Dpol region. Subsequent pairwise comparison of sequences recorded their VP1 sequence divergence (evolutionary separation) and whether the two variants remained clustered in 3Dpol. The likelihood of recombination (i.e. separate grouping of two types in the 3Dpol region) increased steadily with VP1 sequence divergence (Fig. 2a), indicative of time-related recombination comparable to that observed for human enterovirus species A and B sequences (HEV-A, -B; Simmonds & Welch, 2006). Although different HPeV isolates were obtained from different geographical locations and over different collection periods, measurement of VP1 divergence provided an independent measure of their temporal separation from each other, and thus provided a robust comparator with recombination frequency. Furthermore, this independent measure of divergence time allowed us to compare the dynamics of recombination directly with those of HEV-A and -B.

Comparison of the dynamics of recombination of HPeV was carried out by parallel investigation of VP1 sequence divergence and recombination frequency in datasets of HEV-A and -B sequences (Simmonds & Welch, 2006) containing newly published, full-length sequences from 2006–2007. A full listing of the HEV sequences used is available as Supplementary Table S1 in JGV Online. It was not possible to perform a parallel analysis of recombination frequency in HEV-C or HEV-D, because of a lack of published complete-genome sequences of epidemiologically independent isolates (although there are several available
HEV-C sequences, many are from vaccine-derived poliovirus strains). The dynamics of recombination for HPeV variants were remarkably similar to those of the larger HEV-B sequence dataset (Fig. 2a, black and white bars). This similarity extended to a second comparison of recombination frequency and differences in isolation dates (Fig. 2b), an alternative measure of temporal separation of isolates, but one that does not take geographical separation into account.
Consistent with previous findings, the time course (measured by both VP1 divergence and isolation-time differences) of recombination of HEV-A variants was substantially slower than that of HEV-B. Separate analysis of specific HPeV and HEV types showed a similar pattern of increase in recombination when types were analyzed within a species (data not shown).

Despite the detection of frequent recombination events in parechoviruses, recombination was never observed among HPeV3 sequences (Fig. 1). Although the majority of the HPeV3 isolates were isolated within the same year (see Supplementary Table S1, available in JGV Online) and clustered tightly together, a measurable proportion of recombination was detected among HPeV1, 4, 5 and 6 types that were similarly divergent in VP1 (approx. one-fifth of pairwise comparisons where VP1 divergence <0.025 showed recombination). In addition, 50% of non-HPeV3 variants isolated in the same year were recombinant, compared with a frequency of zero for HPeV3 (Figs 1 and 2). These observations suggest possible biological constraints that limit HPeV3 recombination events. In HEVs and other picornavirus genera, the high degree of sequence divergence between species in the non-structural region previously appeared to be the main factor limiting inter-species recombination (Simmonds, 2006). However, analysis of all available full-length sequences (n=11; Fig. 3) showed that HPeV3 was similarly divergent from other parechovirus types (green line) as the latter were from each other (red line). The only exception was the slightly greater sequence divergence between HPeV3 and other types at the C terminus of VP1. This very local region of greater divergence corresponds to the part of the HPeV3 VP1 sequence where the RGD integrin-binding sequence is absent. As the RGD motif in other HPeV types was found to be critical for replication (Boonyakiat et al., 2001), its absence in HPeV3 suggests the use of a different (non-integrin) receptor for entry. Different receptor use can result in a change in cellular tropism and might account for different clinical outcomes observed in HPeV infections, such as more severe disease and central nervous system involvement (Benschop et al., 2006a; Boivin et al., 2005; Ito et al., 2004). Infection of different cell types in vivo might additionally reduce the opportunity for recombination between HPeV3 and other HPeV types to occur, and therefore potentially account for the failure to detect such recombinants in our genetic survey.

Scans of sequence variability across parechovirus genomes revealed much greater amino acid sequence variability and higher $d_N/d_S$ ratios within the structural gene region (Fig. 3). Through calculation of a segregation score in the program TreeOrder Scan in the SIMMONICS sequence editor package (version 1.6; http://www2.warwick.ac.uk/fac/sci/bio/research/devans/bioinformatics/simmonics/), grouping by virus type was observed throughout the structural region, a zone strictly demarcated by the 5'-UTR/VP0 and S/NS (VP1/2A) gene boundaries, the latter also being characterized by a dramatic decline in amino acid sequence variability and $d_N/d_S$ ratio. These observations for HPeV show striking similarities to the pattern of sequence diversity and recombination frequency in other picornaviruses and other mammalian, non-enveloped, positive-stranded RNA viruses (Simmonds, 2006). As a unifying hypothesis for these disparate observations, the dramatically higher $d_N/d_S$ ratio in the structural gene region, combined with its much higher amino acid sequence divergence compared with the rest of the genome, provides further evidence for positive selection operating on the exposed outer surface of the virus. Immune-mediated selection may drive changes in its antigenicity and, combined with changes in receptor use (such as the deletion of the RGD
motif in HPeV), ultimately generate new serotypes refractory to immunity to previously encountered HPeV variants. As proposed previously (Simmonds, 2006), the resulting high amino acid sequence divergence and biological incompatibility may be the key factor, thus preventing the occurrence of recombination in the structural gene region. A second, smaller increase in amino acid diversity and \( d_N/d_S \) ratios was also observed within the 3AB region, although whether this reflects positive selection or less constraint on amino acid sequence change for these amino acids, and for those in the homologous region in HEVs, is unknown. Patterns of amino acid sequence divergence, location of breakpoints and codon usage in parechoviruses revealed in this and previous studies (Simmonds, 2006) thus closely mirror those of other picornavirus groups in which extensive recombination between serotypes has been documented.

This is the first systematic survey of recombination frequencies and temporal dynamics in parechoviruses, and has generated comparative sequence data of VP1 and 3Dpol regions from several HPeV isolates from three different geographical locations. Although there is a limited number of HPeV isolates characterized genetically to date, the limited HPeV dataset used showed a specific evolutionary trend that is also found in other picornaviruses. Recombination may play a major role in the evolution of this virus genus, and was found to occur with similarly rapid temporal dynamics as HEVs. However, more full-length data are needed to study recombination within HPeVs further. The data presented provide further knowledge for studying the molecular evolution and epidemiology of HPeVs and a basis for \textit{in vitro} pathogenesis studies, particularly between HPeV3 and other HPeV types.

**References**


